Identification of the innate normal tissue specific genes and acquired tumor specific genes in determining the tumor transcriptional profiles

Haiwei Wang (hwwang@sibs.ac.cn)
Fujian Provincial Maternity and Children's Hospital https://orcid.org/0000-0002-9675-4039
Xinrui Wang
Fujian Provincial Maternity and Children' Hospital
Liangpu Xu
Fujian Provincial Maternity and Children' Hospital
Ji Zhang
Fujian Provincial Maternity and Children' Hospital
Hua Cao
Fujian Provincial Maternity and Children' Hospital

Research article

Keywords: TCGA, normal tissue specific genes, tumor acquired specific genes, DNA methylation, genomic aberrations.

Posted Date: November 19th, 2019

DOI: https://doi.org/10.21203/rs.2.17488/v1

Abstract

Background: For a specific cancer type, the transcriptional profile is determined by the combination of innate transcriptional features of the original normal tissue and the acquired transcriptional characteristics mediated by genomic and epigenetic aberrations in the tumor development. However, the classification of innate normal tissue specific genes and acquired tumor specific genes is not studied in a pan-cancer manner. Methods: The innate and acquired gene expression profiles in each tumor type were studied using The Cancer Genome Atlas (TCGA) RNA-seq dataset. The prognostic effects of the tumor acquired genes were determined by “survival” package in R software. The methylation of the tumor acquired genes was delineated using TCGA HumanMethylation450 microarray data. Results: 90% liver hepatocellular carcinoma (LIHC) specific genes are derived from innate normal liver specific genes. On the contrary, 90.3% kidney clear cell carcinoma (KIRC) specific genes and 90.9 % lung squamous cell carcinoma (LUSC) specific genes are acquired in the tumor developmental progress. The innate normal tissue specific genes are down regulated in tumor tissues, while, the tumor acquired specific genes are up regulated in the tumor tissues. The innate normal tissue specific genes and the tumors acquired specific genes are both associated with the tumor overall survival in some tumor types. The hyper-DNA methylation of normal tissue specific genes is contributing to the inhibition of normal tissue specific genes expression in cancer cells. And the tumor acquired specific genes are activated by hypo-DNA methylation and genomic aberrations. Conclusions: Our results provide descriptions of the specific transcriptional features across cancer types and suggest that the tumor acquired specific genes are potential targets for anti-cancer therapy.

Background

Cancer is usually classified by the original normal tissue where the tumor cells are derived from. Due to the different cell original patterns, each tumor type has a very distinctive and unique transcriptional feature [1-5]. However, the original tissue expression characteristics, only influence, but not fully determine the tumor classifications [6]. Those observations highlight the contributions of genetic and epigenetic changes in determining the distinctive transcriptional profiles of tumor cells.

Genetic changes are including genomic rearrangements, gene amplifications or deletions, and specific gene mutations. The genomic aberrations are highly important to tumor therapy response and tumor overall survival [7]. Epigenetic changes such as DNA methylation and chromatin modification are also critical to cancer development and progress [8-10]. Those genetic and epigenetic changes are finally reflected to transcriptome, including mRNA, microRNA and lncRNAs deregulations and proteome, including protein expression and modification alterations in cancer cells.

So, for a specific cancer type, the ultimate transcriptional profiles are determined by the combination of innate transcriptional features of the original normal tissue and acquired transcriptional characteristics by genomic and epigenetic aberrations. However, which factor is more important to determine the different transcriptional features across tumor types is no clear. And due to the variation in cancer driver

alterations among different tumor types, the acquired tumor transcriptional characteristics may dramatically be different.

With the advances of TCGA project, the genetic and epigenetic changes as well as the transcriptional alterations of each tumor type and across tumor types are well illustrated [11, 12]. Moreover, some normal tissue expression and DNA methylation data is deposited in TCGA project [13, 14]. And all the data is open-accessed, thus providing us great knowledge to address how the genetic, epigenetic changes and innate transcriptional difference of normal tissues influence the tumor transcriptional features across cancer types.

Here, the malignant and normal tissue specific genes are identified across 14 tumor types. The normal tissue specific genes are down regulated in tumor tissues. After overlapping normal and tumor specific transcriptional feature, we find that the innate transcriptional profiles of normal liver determine the transcriptional features of LIHC. However, KIRC and LUSC specific genes are mainly new acquired by genomic and epigenetic aberrations. Particularly, KIRC acquired specific genes are activated by hypo-DNA methylation, while, LUSC acquired specific genes are activated by DNA amplifications. Tumor acquired specific genes in other tumor types are also studied. So, our analysis provide molecular understandings of how the tumor transcriptional profiles are determined by the combination of innate transcriptional feature of normal tissue and acquired transcriptional characteristics by genomic and epigenetic aberrations.

Methods

Data collection

Gene expression profiles across cancer types were analyzed using RNA-seq data (TCGA HiSeqV2 data). The DNA methylation profiles were analyzed through HumanMethylation450 microarray data. All the datasets were downloaded from the TCGA hub (https://tcga.xenahubs.net).

Identification of the normal and malignant tissue specific genes

The average count (log2) of each gene in the various malignant tissue samples was calculated. For a gene with count > 512 (log2 count > 9) and 1.5 fold higher than any other tissues was classified into malignant tissue specific gene. Same selection threshold was used to identify the normal tissue specific genes.

Heatmap

Heatmaps were created by R software “pheatmap” package. The “pheatmap” package was available in bioconductor. The clustering scale was determined by "average" method.

Survival analysis

Kaplan-Meier estimator from “survival” package in the R statistics software was applied to identify the association between tumor acquired tissue specific genes and tumor overall survival. The “survival” package and the basic usage were downloaded from bioconductor. Log-rank P value was determined.

Analysis the genomic alteration

LUSC and LIHC acquired specific genes amplifications were downloaded from cbioportal (http://www.cbioportal.org/index.do). The DNA location of genes was annotated according to hg19.

Venny diagram

The venny diagrams of normal and malignant tissue specific genes were generated by VENNY 2.1 (http://bioinfogp.cnb.csic.es/tools/venny/index.html).

Tumor acquired specific core transcription factors network

The networks of tumor acquired specific core transcription factors were created by cytoscape GeneMANIA App. The first degrees of core transcription factor connected genes were demonstrated.

Statistical analysis

The box plots were generated from prims5.0. Statistical analysis was performed using the Student’s t test.

Results

Identification of the malignant tissue specific genes from TCGA dataset.

To identify the tumor tissue specific genes, we evaluated all the tumor samples in the TCGA collection which the RNA-seq data was available. Only samples with corresponding normal samples were used for further studies. The expression level of each gene was calculated. For a gene with log2 count > 9 and 1.5 fold higher than any other tumor tissues was classified into tissue specific gene. This resulted in a final set of 2738 tumor tissue specific genes from 645 corresponding tumor samples across 14 different cancer types, including bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), head and neck squamous cell carcinoma (HNSC), kidney chromophobe (KICH), KIRC, kidney papillary cell carcinoma (KIRP), LIHC, lung adenocarcinoma (LUAD), LUSC, prostate adenocarcinoma (PRAD), stomach adenocarcinoma (STAD) and thyroid carcinoma (THCA) (Fig. 1). The number of tumor specific genes was also varied significantly from different tumor types (Fig. 1). There were 579 LIHC specific genes, while, there were only 36 STAD specific genes and 36 BLCA specific genes. LUSC had the least 11 specific genes (Fig. 1).

Although, those tumor specific genes were highly expressed in corresponding malignant tissues, we found that COAD specific genes and ESCA were also highly expressed in STAD tissues (Fig. 1). Those results further highlighted the similar functions and tissue origin of colon, esophagus and stomach [15].

Another interesting finding was that HNSC specific genes were also highly expressed in LUSC tissues (Fig. 1). This phenomenon will be further illustrated.

Identification of the normal tissue specific genes from TCGA dataset.

Using the same selective strategies, normal tissue specific genes from TCGA dataset were identified. Totally 645 normal samples from 11 different tissue types, including bladder, breast, colon, esophagus, head, neck, kidney, liver, lung, prostate, stomach and thyroid were studied (Fig. 2). Normal kidney samples were combined from KICH, KIRP and KIRC datasets. Normal lung samples were combined from LUAD and LUSC datasets. This resulted in 3368 normal tissue specific genes. As illustrated in the heatmap presentation (Fig. 2), those genes were highly expressed in corresponding tissues. The number of tissue specific genes was also varied significantly from different tissue types. For instant, there were 1089 liver specific genes, while, there were only 13 stomach specific genes and 67 bladder specific genes (Fig. 2). Notably, because of the functional similarity, stomach shared similar transcriptional features with the gastrointestinal tract, head and neck tissues [15] (Fig. 2).

The overlapping between normal tissue specific genes and malignant tissue specific genes.

Therefore, we obtained both the normal tissue specific genes and corresponding tumor specific genes, we then determined the overlapped normal and tumor specific genes. The venny diagrams depicted the common and unique genes between normal and tumor specific genes across 14 tumor types (Fig. 3). For the majority of tumor types, the tumor tissue specific genes were few than normal tissue specific genes (Fig. 3). We suggested that the higher percentage of common genes derived from tumor specific genes indicating more importance of innate transcriptional profiles of normal tissues in determining the transcriptional features of tumor cells.

We found that 90% LIHC specific genes were derived from normal liver specific genes, suggesting the innate transcriptional profiles of normal liver were more important to determine the transcriptional features of LIHC (Fig. 3). 57.38% colonic COAD, 57% HNSC and 67.5% LUAD specific genes were derived from normal colon, head and neck and lung specific genes, respectively (Fig. 3). On the contrary, only 9.7% KIRC specific genes and 9.1% LUSC specific genes were derived from normal kidney and lung specific genes respectively, suggesting the genomic aberrations and DNA methylation were more important to determine the transcriptional features of KIRC and LUSC across tumor types (Fig. 3).

The normal tissue specific genes are decreased in the tumor samples.

Tumor is an abnormal growth of tissue losing the original specialized functions. The collapse of original specialized functions may be induced by the loss of tissue specific genes in tumor cells. So, we tested the expression of tissue specific genes in normal tissues and corresponding tumor tissues. We found that the tissue specific genes were inhibited in BLCA, BRCA, COAD, ESCA, HNSC, KIRC, KIRP, LIHC, LUAD, LUSC and THCA tumor samples (Fig. 4). Particularly, compared to the normal samples, nearly all the colon specific genes were inhibited in COAD, kidney specific genes were inhibited in KIRC and KIRP, liver specific genes

were inhibited in LIHC and lung specific genes were inhibited in LUAD and LUSC (Fig. 4). However, we found that most of the prostate specific genes were not inhibited in PRAD tumor samples (Fig. 4).

The normal tissue specific genes are inhibited in cancer by hyper-DNA methylation.

Next, we tried to determine the mechanisms that induced the decreasing of tissue specific genes in the tumor development. The first clue was DNA methylation. Tissue specific genes were highly controlled by DNA methylation [16]. Compared to the normal samples, the alterations of methylation profile across cancer types were studied in TCGA [13, 14]. Using the normal samples in the TCGA for which the DNA methylation data was available, we showed that the tissue specific genes were with low DNA methylation in corresponding tissues (Fig. 5a).

In the tumor developmental process, the inhibited tissue specific gene expression may be controlled by DNA hyper-methylation. And the increased DNA methylation was controlled by DNA methyltransferase hyper-activity [8-10]. DNMT1 is a DNA methyltransferase and is responsible for the maintaining of the DNA methylation patterns [17]. We found that compared to the normal samples, DNMT1 was up regulated in nearly all the tumor types, except PRAD (Fig. 5b).

The elevated DNMT1 expression in tumor may increase the DNA methylation in the tissue specific genes thus inhibited their expressions. So, the DNA methylation profiles between normal and tumor samples in different tumor types were analyzed. Compared to the normal samples, we found that nearly half percentage of tissue specific genes were with high DNA methylation patterns in BRAC, KIRC, KIRP, LIHC, LUAD and LUSC tumor types (Fig. 6c). Those observations suggested that DNA methylation was partially contributing to the decreasing of tissue specific genes in the tumor development.

The tumor acquired specific genes are increased in the tumor samples.

From above results, we had shown that the normal tissue specific genes were decreased in the tumor samples. We proposed that in the venny diagrams (Fig. 3), the unique normal tissue specific genes were totally lost due to the dedifferentiation process of the tumor development. Although, the common genes were also decreased in the tumor samples, those genes still maintained the tissue specifications in tumor cells. The tumor unique specific genes were new acquired genes which were up regulated by genomic aberrations or mis-regulations of DNA methylation.

To test this proposition, the expressions of tumor acquired specific genes in normal tissues and corresponding tumor tissues were illustrated. We found that most of the tumor acquired specific genes were highly expressed in BLCA, BRCA, COAD, ESCA, HNSC, KICH, KIRC, KIRP, LUAD, LUSC, PRAD and THCA tumor samples than the normal samples (Fig. 6).

The KIRC acquired specific genes are activated by hypo-DNA methylation.

Next, we tried to determine the mechanisms that induced the activation of tumor acquired specific genes in the tumor development. Previously, we had shown that the hyper-DNA methylation was partially

determining the decreasing of normal tissue specific genes in the tumor (Fig. 5c). Contrast to the normal samples, we found that some of the tumor acquired specific genes were with hypo-DNA methylation patterns in BLCA, BRCA, COAD, HNSC, KIRC, LIHC, LUAD and THCA tumor types (Fig. 7a). Particularly, in KIRC tumor type, more than 80% tumor acquired specific genes were with hypo-DNA methylation in tumor samples (Fig. 7a). Those observations suggested that DNA methylation was partially contributing to the activation of tumor acquired specific genes in the tumor cells.

The LUSC and LIHC acquired specific genes are activated by DNA amplifications.

Another factor determining to the activation of tumor acquired specific genes in tumor cells was genomic aberrations, particular DNA amplification. We found that the LUSC acquired specific genes were with significant DNA amplifications (Fig. 7b). LUSC acquired specific gene SOX2 was amplified in 40% LUSC patients, playing important roles in LUSC development (Fig. 7b). LUSC acquired specific genes YEATS2, ZYF639, MYNN, RFC4 and RPL39L were also amplified in more than 30% LUSC patients (Fig. 7b). Interestingly, SOX2, YEATS2, ZYF639, MYNN, RFC4 and RPL39L were all located in 3q11.1 DNA region, suggested that the LUSC acquired genes were in the same amplified gene cluster (Fig. 7b).

There were 57 LIHC acquired specific genes. Among them, we found 18 genes were amplified in more than 10% LIHC patients (Fig. 7c). And those 18 genes were located on two DNA regions, 8q11.21 and 1q12 (Fig. 7c). The DNA amplifications of acquired specific genes in other tumor types were also studied and not significant DNA amplifications were observed.

Acquired of head and neck normal specific genes in LUSC development is mediated by SOX2 amplification.

SOX2 plays important roles in embryonic development, maintaining pluripotent stem cells identity and cell differentiation [18-20]. Previously, it had reported that SOX2 was associated with increased cancer aggressiveness [21, 22] and therapy resistance [23]. We found that SOX2 amplified and acquired in LUSC (Fig. 7b) and involved in regulation of LUSC specific genes. We also showed that SOX2 was particular expressed in head and neck tissue and SOX2 involved regulatory networks was important to maintaining head and neck functions (Fig. 7d).

LUAD and LUSC were both derived from normal lung tissue. Further demonstration suggested that, in LUSC, the original normal lung expression profiles were totally lost. Instead, LUSC was highly expressed head and neck specific genes (Fig. 7d). Those observations provided some explanations that HNSC specific genes were also highly expressed in LUSC tissues (Fig. 1).

The normal tissue specific genes and the tumor acquired specific gene expression are associated with the tumor outcomes.

At last, we tested whether the down regulation of tissue specific genes and the up regulation of the acquired specific genes were associated with the tumor progress. Cohort of tumor expression data with clinical overall survival in TCGA dataset was studied. The kaplan-meier survival analysis showed that the

tissue specific genes distinguished a cluster of patients with high probability of overall survival in BLCA (P=2e-04), HNSC (P=0.002), LIHC (P=0.03), KIRC (P=4e-05) and KIRP (P=0.004) (Fig. 8a). However, the tissue specific genes were not associated with the overall survival in LUAD, BRCA, COAD and STAD tumor types (Fig. 8a).

The kaplan-meier survival analysis also showed that the acquired specific genes distinguished a cluster of patients with low probability of overall survival in KICH (P=0.07), KIRC (P=0.008) and KIRP (P=0.03) (Fig. 8b). However, the tumor acquired specific genes were not associated with the overall survival in other tumor types.

Discussion

Here, normal and malignant tissue specific genes are identified from TCGA dataset (Fig. 1 and 2) and those genes are further studied in tumor samples. We suggest that the down regulation of normal tissue specific genes in cancer is contributing to the collapse of normal tissue functions (Fig. 4). And the inhibitions of normal tissue specific genes in cancer are caused by epigenetic hyper-DNA methylation (Fig. 5). However, how the genomic alterations induce the down regulation of tissue specific genes and collapse of normal tissue functions are still not known, and difficult to be addressed from TCGA dataset. We did try to determine whether the DNA deletion was an inner mechanism to induce the inhibition of tissue specific genes in cancer. However, almost all the tissue specific genes showed no DNA deletion. Whether the cancer drive mutations influence the tissue specific gene expression is even harder to determine, because of the different cancer drive mutations across tumor types. TP53 is the most common drive mutation [24]. And loss of TP53 induces multiple types of cancer in TP53 knockout mice [25]. Next, we try to test the tissue specific gene expression in TP53 wild type and knockout mice to determine whether loss of TP53 functions inhibits the tissue specific gene expression.

Tumor development is a dedifferentiation process [26, 27]. The normal cells lose the original specialized functions and dedifferentiated to a relative primary state. Also in the tumor development process, cancer cells acquire some tumor hallmarks because of the accumulation of genetic and epigenetic alterations [28, 29]. However, in responding to the genomic variations in the tumor cells, each tissue has different abilities to maintain its original characteristics (Fig. 3). Compared to other tumor types, LIHC has the most ability to maintain its original characteristics. Although, our results showed the decreasing of liver specific genes in LIHC (Fig. 4), the remaining transcriptional features still distinguish LIHC from other tumor types. KIRC and LUSC have the least ability to maintain its original characteristics (Fig. 3). LUSC acquires partial head and neck transcriptional features (Fig. 7d). The acquired LUSC specific genes are activated by DNA amplifications, particular LUSC acquired specific gene SOX2 are amplified in 40% LUSC patients (Fig. 7b). The acquired KIRC specific genes are activated by hypo-DAN methylation. More than 80% KIRC specific genes are with low DNA methylation (Fig. 7a).

The BLCA, BRCA, COAD, HNSC, LIHC, LUAD and THCA acquired specific genes are also partially determined by hypo-DNA methylation or DNA amplifications (Fig. 7a).. However, the inner mechanisms

of ESCA, KICH, PRAD and STAD acquired specific genes are not yet determined. We find no hypo-DNA methylation or DNA amplifications in ESCA, KICH, PRAD and STAD acquired specific genes. So, we hypothesize that ESCA, KICH, PRAD and STAD acquired specific genes are indirectly activated by specific genetic changes. Like TP53 signaling pathway is mutant in more than 50% ESCA, KICH and STAD patients [7]. MYC mediated transcription network is amplified in ESCA and STAD [30]. Next, we will use experiments to validate this hypothesis.

In our paper, we analyzed the normal tissue specific genes across cancer types, and provided molecular understandings of the losing of original tissue specialized functions in tumor. We also identify the tumor unique gene expression profiles and the distinctive tumor acquired specific genes. So, our data provide molecular understandings of how the tumor transcriptional profiles are determined by the combination of innate transcriptional feature of normal tissue and acquired transcriptional characteristics by genomic and epigenetic aberrations. Further experimental studies and validations of those tumor acquired specific genes will provide great understanding of tumor development and tumor overall survival. Targeting those tumor acquired specific genes may be efficient in tumor treatment and without significant side effects.

Conclusions

Our results provide descriptions of the specific transcriptional features across cancer types and suggested that the tumor acquired specific genes are potential targets for anti-cancer therapy.

Abbreviations

TCGA: The Cancer Genome Atlas; LIHC: liver hepatocellular carcinoma; KIRC: kidney clear cell carcinoma; LUSC: lung squamous cell carcinoma; BLCA: bladder urothelial carcinoma; BRCA: breast invasive carcinoma; COAD: colon adenocarcinoma; ESCA: esophageal carcinoma; HNSC: head and neck squamous cell carcinoma; KICH: kidney chromophobe; KIRP: kidney papillary cell carcinoma; LUAD: lung adenocarcinoma; PRAD: prostate adenocarcinoma; STAD: stomach adenocarcinoma; THCA: thyroid carcinoma;

Declarations

Ethics approval and consent to participate

None of the materials are required ethical approval for their use.

Consent for publication

Not applicable.

Availability of data and material

All the data and software used in this paper are available as mentioned in Methods.

Competing interests

The authors declare no conflicts of interests.

Funding

This study was supported by grants from Fujian Provincial Maternity and Children's Hospital (No.YCXB 18-10 and YCXM 19-04). This study was also supported by National Natural Science Foundation (No.81370655), and by Distinguished Professorship (JZ) from Shanghai Jiao Tong University School of Medicine, Shanghai Jiao Tong University.

Authors' contributions

HW.W designed and performed data analysis. XR.W and LP.X helped with the data analysis. HW.W, HC and JZ wrote the manuscript. HC and JZ reviewed the manuscript and supervised the work.

Acknowledgements

All the results shown here are based upon data generated by the TCGA Research Network (http://cancergenome.nih.gov/).

References

1. Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhori G, Benfeitas R, Arif M, Liu Z, Edfors F *et al*: **A pathology atlas of the human cancer transcriptome**. *Science* 2017, **357**(6352).
2. Yan X, Hu Z, Feng Y, Hu X, Yuan J, Zhao SD, Zhang Y, Yang L, Shan W, He Q *et al*: **Comprehensive Genomic Characterization of Long Non-coding RNAs across Human Cancers**. *Cancer Cell* 2015, **28**(4):529-540.
3. Chiu HS, Somvanshi S, Patel E, Chen TW, Singh VP, Zorman B, Patil SL, Pan Y, Chatterjee SS, Sood AK *et al*: **Pan-Cancer Analysis of lncRNA Regulation Supports Their Targeting of Cancer Genes in Each Tumor Context**. *Cell Rep* 2018, **23**(1):297-312 e212.
4. Chu A, Robertson G, Brooks D, Mungall AJ, Birol I, Coope R, Ma Y, Jones S, Marra MA: **Large-scale profiling of microRNAs for The Cancer Genome Atlas**. *Nucleic Acids Res* 2015, **44**(1):e3.
5. Dhawan A, Scott JG, Harris AL, Buffa FM: **Pan-cancer characterisation of microRNA across cancer hallmarks reveals microRNA-mediated downregulation of tumour suppressors**. *Nat Commun* 2018, **9**(1):5228.
6. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, Shen R, Taylor AM, Cherniack AD, Thorsson V *et al*: **Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer**. *Cell* 2018, **173**(2):291-304 e296.
7. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, Colaprico A, Wendl MC, Kim J, Reardon B *et al*: **Comprehensive Characterization of Cancer Driver Genes and Mutations**. *Cell*

2018, **173**(2):371-385 e318.

8. Jones PA, Baylin SB: **The epigenomics of cancer**. *Cell* 2007, **128**(4):683-692.
9. Jones PA, Baylin SB: **The fundamental role of epigenetic events in cancer**. *Nat Rev Genet* 2002, **3**(6):415-428.
10. Esteller M: **Cancer epigenomics: DNA methylomes and histone-modification maps**. *Nat Rev Genet* 2007, **8**(4):286-298.
11. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM: **The Cancer Genome Atlas Pan-Cancer analysis project**. *Nat Genet* 2013, **45**(10):1113-1120.
12. Hutter C, Zenklusen JC: **The Cancer Genome Atlas: Creating Lasting Value beyond Its Data**. *Cell* 2018, **173**(2):283-285.
13. Weisenberger DJ: **Characterizing DNA methylation alterations from The Cancer Genome Atlas**. *J Clin Invest* 2016, **124**(1):17-23.
14. Saghafinia S, Mina M, Riggi N, Hanahan D, Ciriello G: **Pan-Cancer Landscape of Aberrant DNA Methylation across Human Tumors**. *Cell Rep* 2018, **25**(4):1066-1080 e1068.
15. Liu Y, Sethi NS, Hinoue T, Schneider BG, Cherniack AD, Sanchez-Vega F, Seoane JA, Farshidfar F, Bowlby R, Islam M *et al*: **Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas**. *Cancer Cell* 2018, **33**(4):721-735 e728.
16. Sun W, Bunn P, Jin C, Little P, Zhabotynsky V, Perou CM, Hayes DN, Chen M, Lin DY: **The association between copy number aberration, DNA methylation and gene expression in tumor samples**. *Nucleic Acids Res* 2018, **46**(6):3009-3018.
17. Jeltsch A, Jurkowska RZ: **New concepts in DNA methylation**. *Trends Biochem Sci* 2014, **39**(7):310-318.
18. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA: **Master transcription factors and mediator establish super-enhancers at key cell identity genes**. *Cell* 2013, **153**(2):307-319.
19. White MD, Angiolini JF, Alvarez YD, Kaur G, Zhao ZW, Mocskos E, Bruno L, Bissiere S, Levi V, Plachta N: **Long-Lived Binding of Sox2 to DNA Predicts Cell Fate in the Four-Cell Mouse Embryo**. *Cell* 2016, **165**(1):75-87.
20. Goolam M, Scialdone A, Graham SJL, Macaulay IC, Jedrusik A, Hupalowska A, Voet T, Marioni JC, Zernicka-Goetz M: **Heterogeneity in Oct4 and Sox2 Targets Biases Cell Fate in 4-Cell Mouse Embryos**. *Cell* 2016, **165**(1):61-74.
21. Mu P, Zhang Z, Benelli M, Karthaus WR, Hoover E, Chen CC, Wongvipat J, Ku SY, Gao D, Cao Z *et al*: **SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer**. *Science* 2017, **355**(6320):84-88.
22. Maurizi G, Verma N, Gadi A, Mansukhani A, Basilico C: **Sox2 is required for tumor development and cancer cell proliferation in osteosarcoma**. *Oncogene* 2018, **37**(33):4626-4632.

23. Piva M, Domenici G, Iriondo O, Rabano M, Simoes BM, Comaills V, Barredo I, Lopez-Ruiz JA, Zabalza I, Kypta R *et al*: **Sox2 promotes tamoxifen resistance in breast cancer cells**. *EMBO Mol Med* 2014, **6**(1):66-79.
24. Bouaoun L, Sonkin D, Ardin M, Hollstein M, Byrnes G, Zavadil J, Olivier M: **TP53 Variations in Human Cancers: New Lessons from the IARC TP53 Database and Genomics Data**. *Hum Mutat* 2016, **37**(9):865-876.
25. Donehower LA, Lozano G: **20 years studying p53 functions in genetically engineered mice**. *Nat Rev Cancer* 2009, **9**(11):831-841.
26. Friedmann-Morvinski D, Verma IM: **Dedifferentiation and reprogramming: origins of cancer stem cells**. *EMBO Rep* 2014, **15**(3):244-253.
27. Slack JM: **Metaplasia and transdifferentiation: from pure biology to the clinic**. *Nat Rev Mol Cell Biol* 2007, **8**(5):369-378.
28. Hanahan D, Weinberg RA: **The hallmarks of cancer**. *Cell* 2000, **100**(1):57-70.
29. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation**. *Cell* 2011, **144**(5):646-674.
30. Schaub FX, Dhankani V, Berger AC, Trivedi M, Richardson AB, Shaw R, Zhao W, Zhang X, Ventura A, Liu Y *et al*: **Pan-cancer Alterations of the MYC Oncogene and Its Proximal Network across the Cancer Genome Atlas**. *Cell Syst* 2018, **6**(3):282-300 e282.

Figures

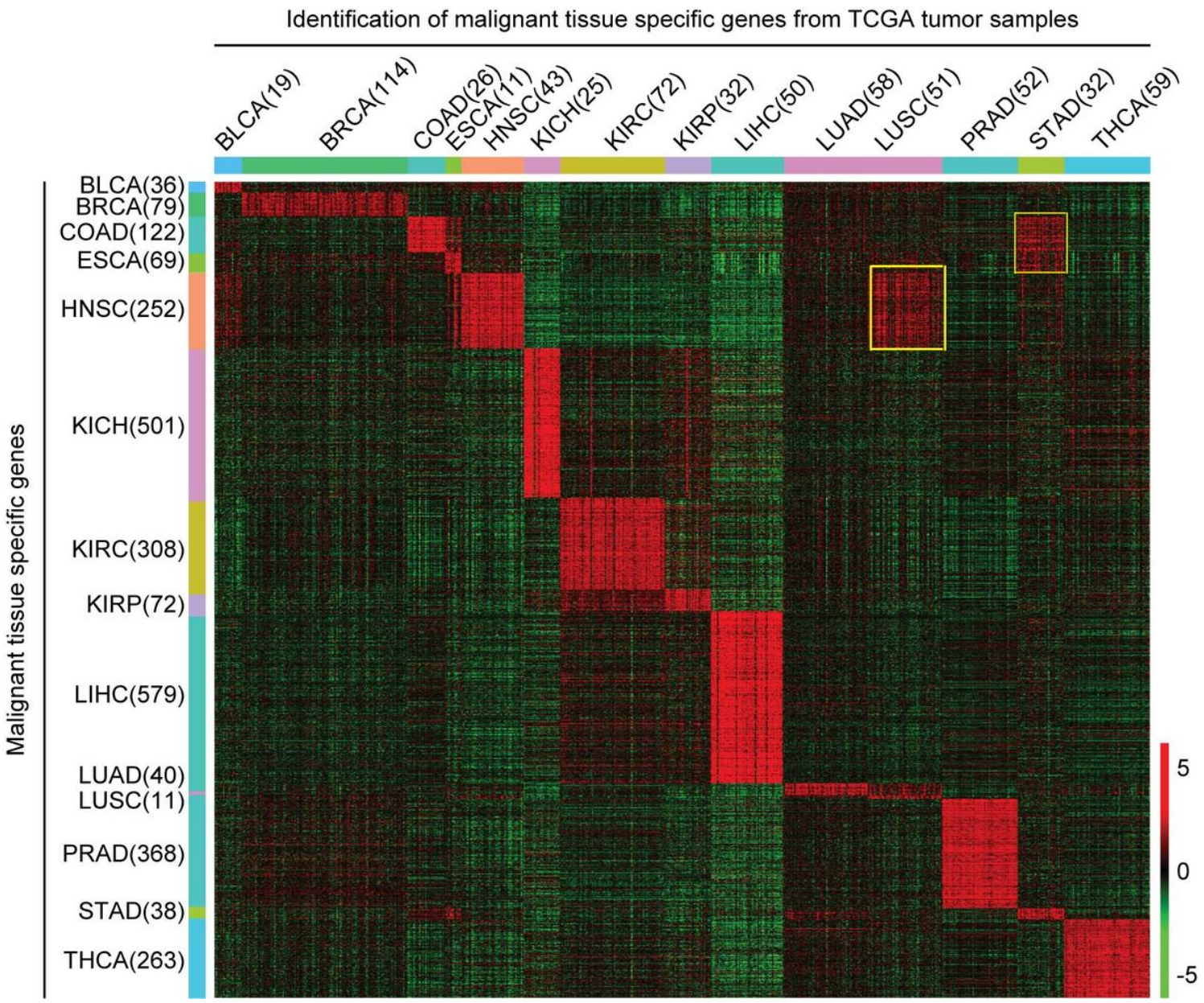

Figure 1

Identification of the malignant tissue specific genes from TCGA dataset. Heatmap showed the malignant tissue specific gene expression in TCGA RNA-seq data. High-regulated (red) and low-regulated (green) genes were demonstrated. The number of malignant tissue sample and malignant tissue specific genes was annotated.

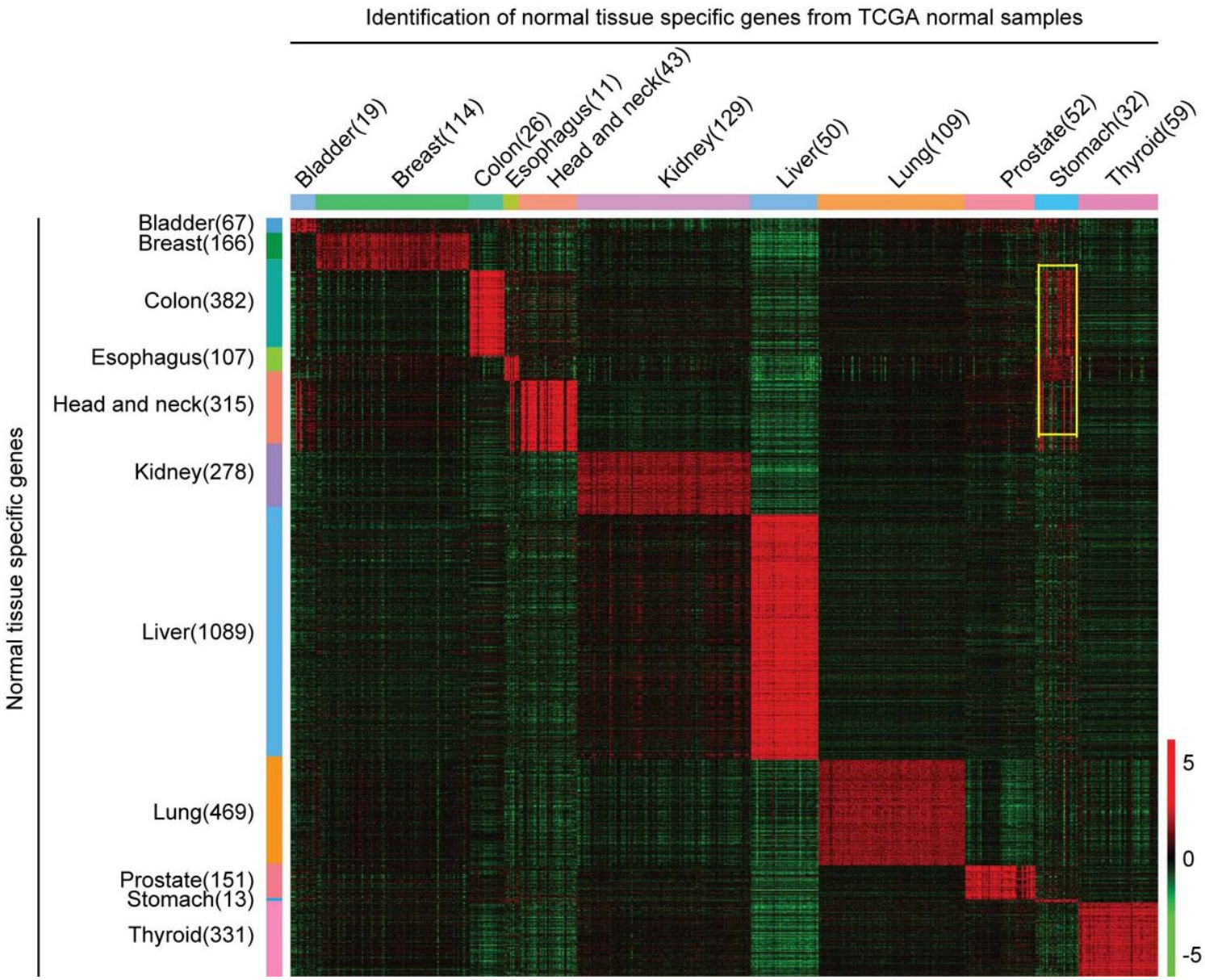

Figure 2

Identification of the normal tissue specific genes from TCGA dataset. Unsupervised clustering heatmap demonstrated the expression of the normal tissue specific genes identified from TCGA dataset. The number of normal tissue sample and normal tissue specific genes was annotated.

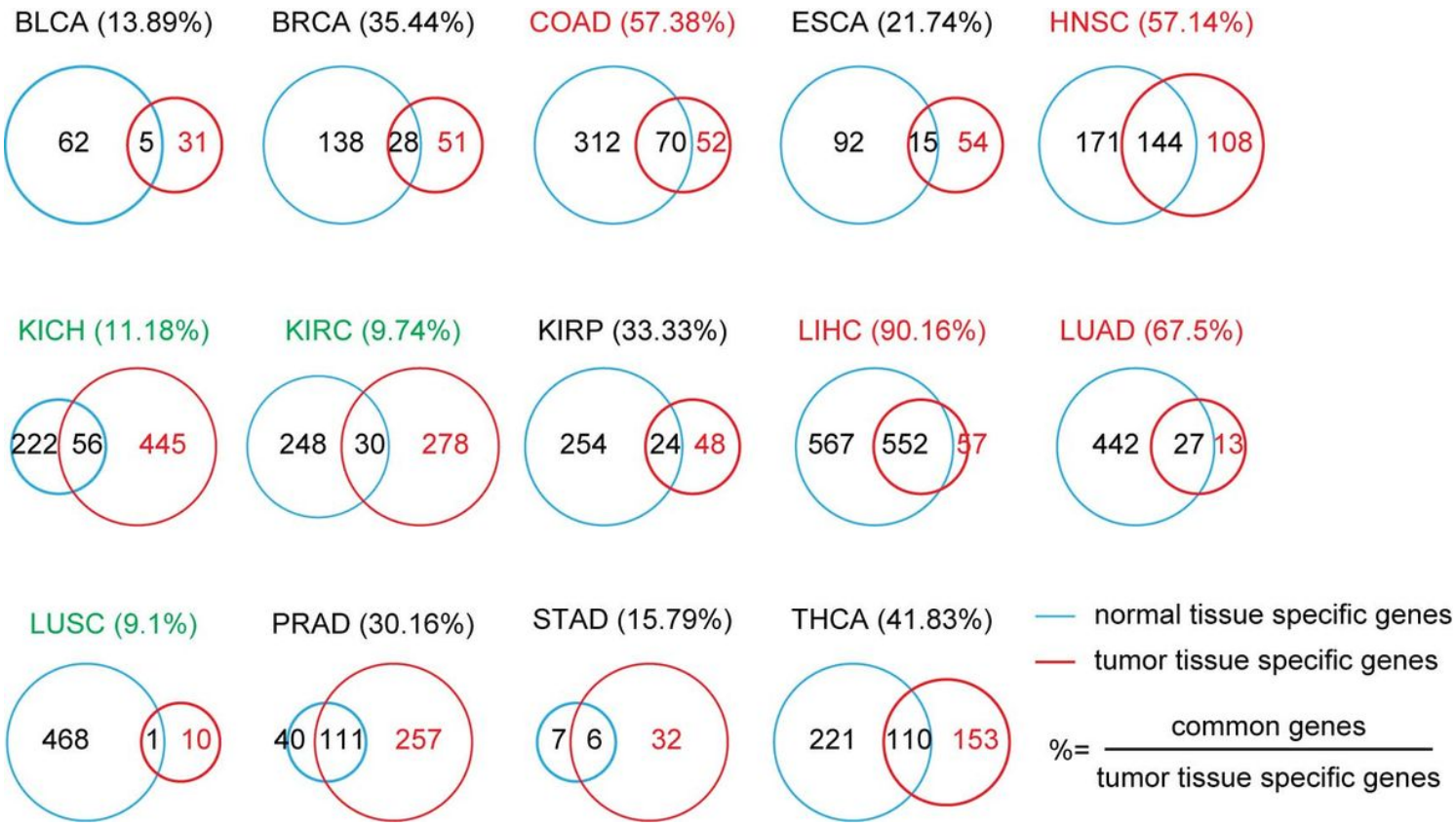

Figure 3

The overlapping between normal tissue specific genes and malignant tissue specific genes. Venny diagrams depicted the common and unique genes between normal and tumor tissue specific genes across 14 tumor types.

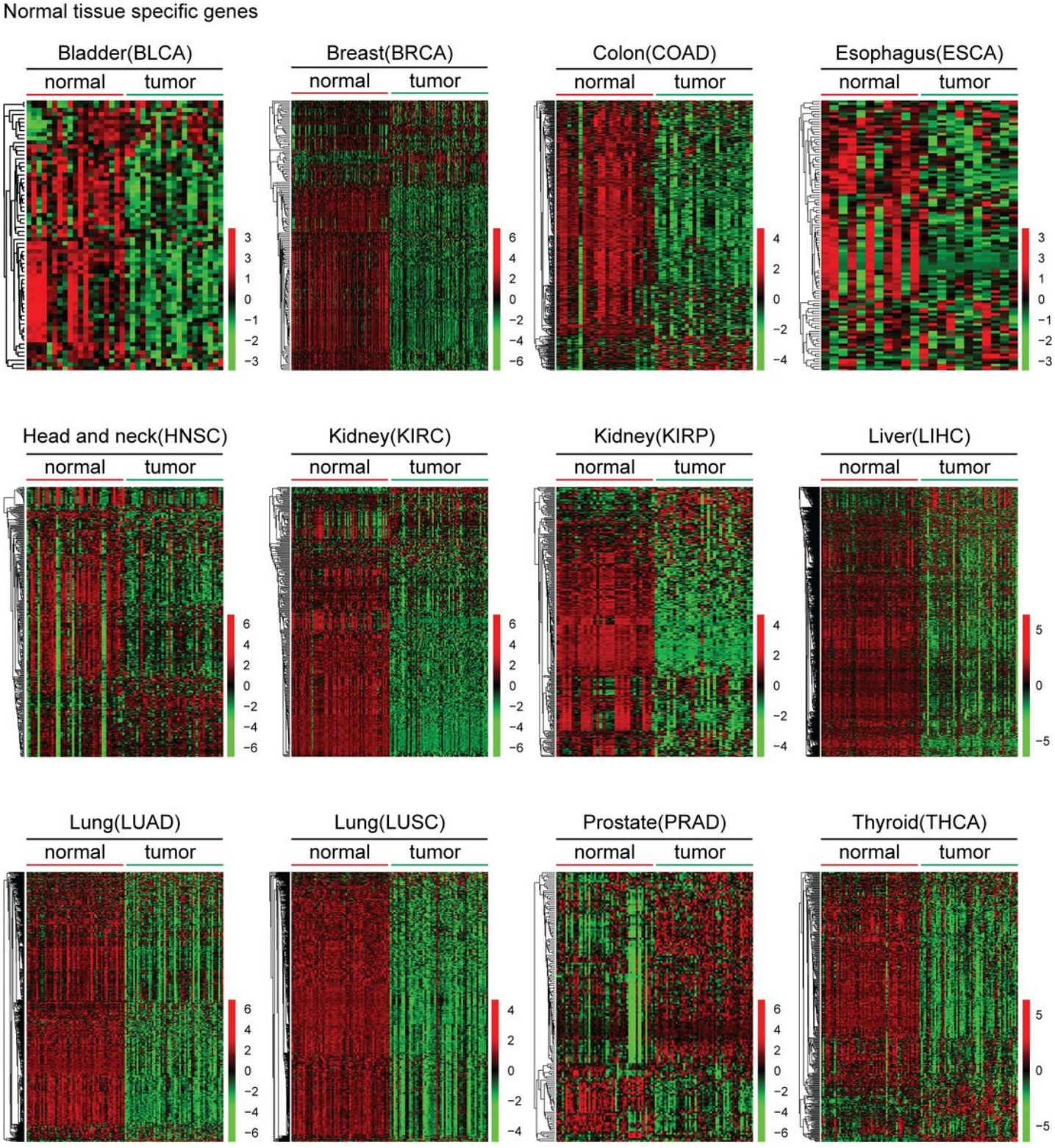

Figure 4

The normal tissue specific genes are decreased in tumor samples. Heatmaps showed the normal tissue specific genes expression in normal and tumor samples from TCGA dataset. Kidney specific genes were demonstrated from KIRC and KIRP tumor samples. Lung specific genes were demonstrated from LUAD and LUSC tumor samples.

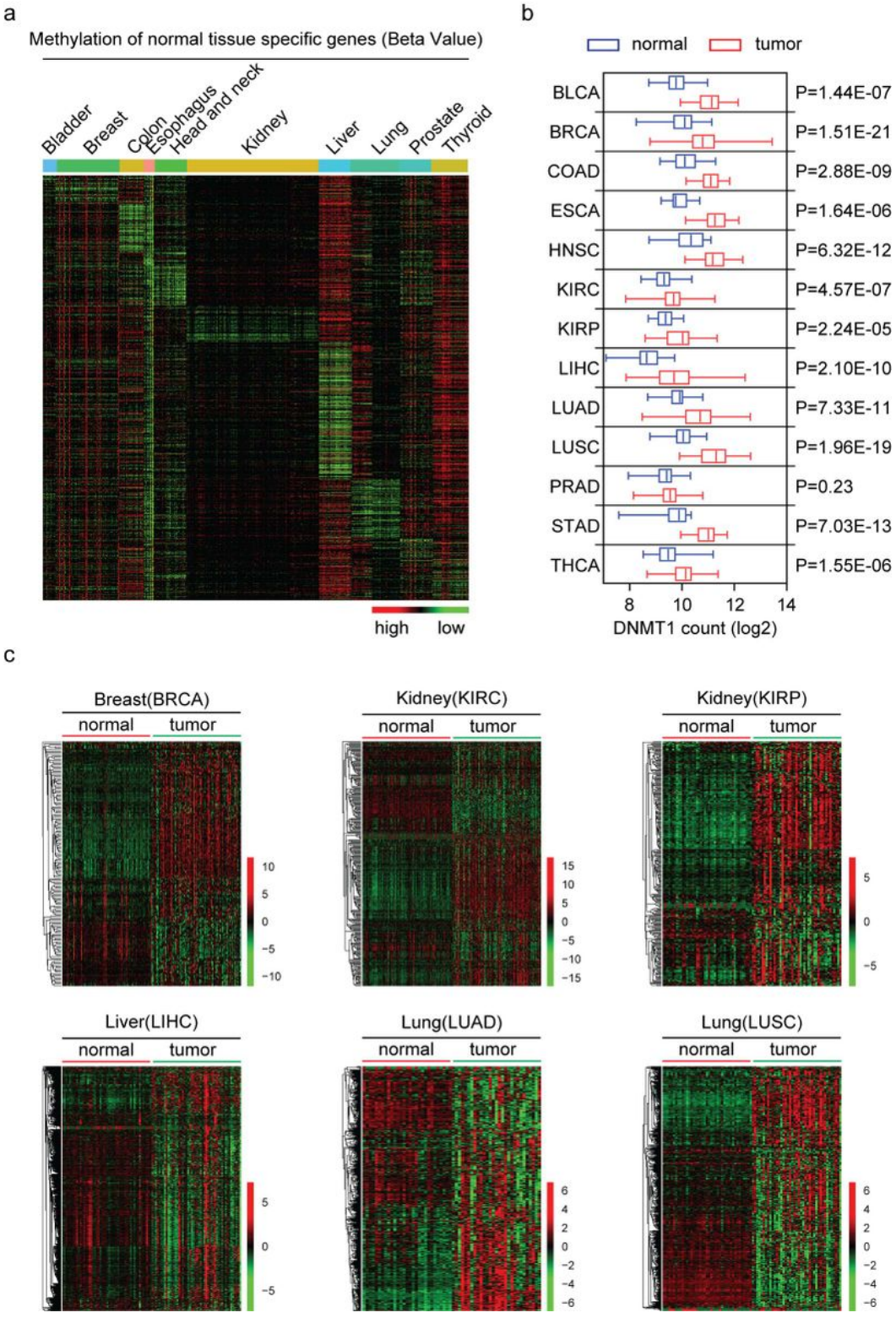

Figure 5

The normal tissue specific genes are inhibited in cancer by hyper-DNA methylation. (a) Heatmap showed the methylation (Beta value) of normal tissue specific genes in TCGA DNA methylation data. Hyper-methylated (red) and hypo-methylated (green) genes were delineated. (b) Box plots showed the DNMT1 expressions (log2 normalization count) in 14 tumor types. P values showed the difference between normal (blue) and tumor (red) samples determined by Student's t test. (c) Heatmaps showed the

methylation (Beta value) of normal tissue specific genes in normal and tumor samples in each tumor type.

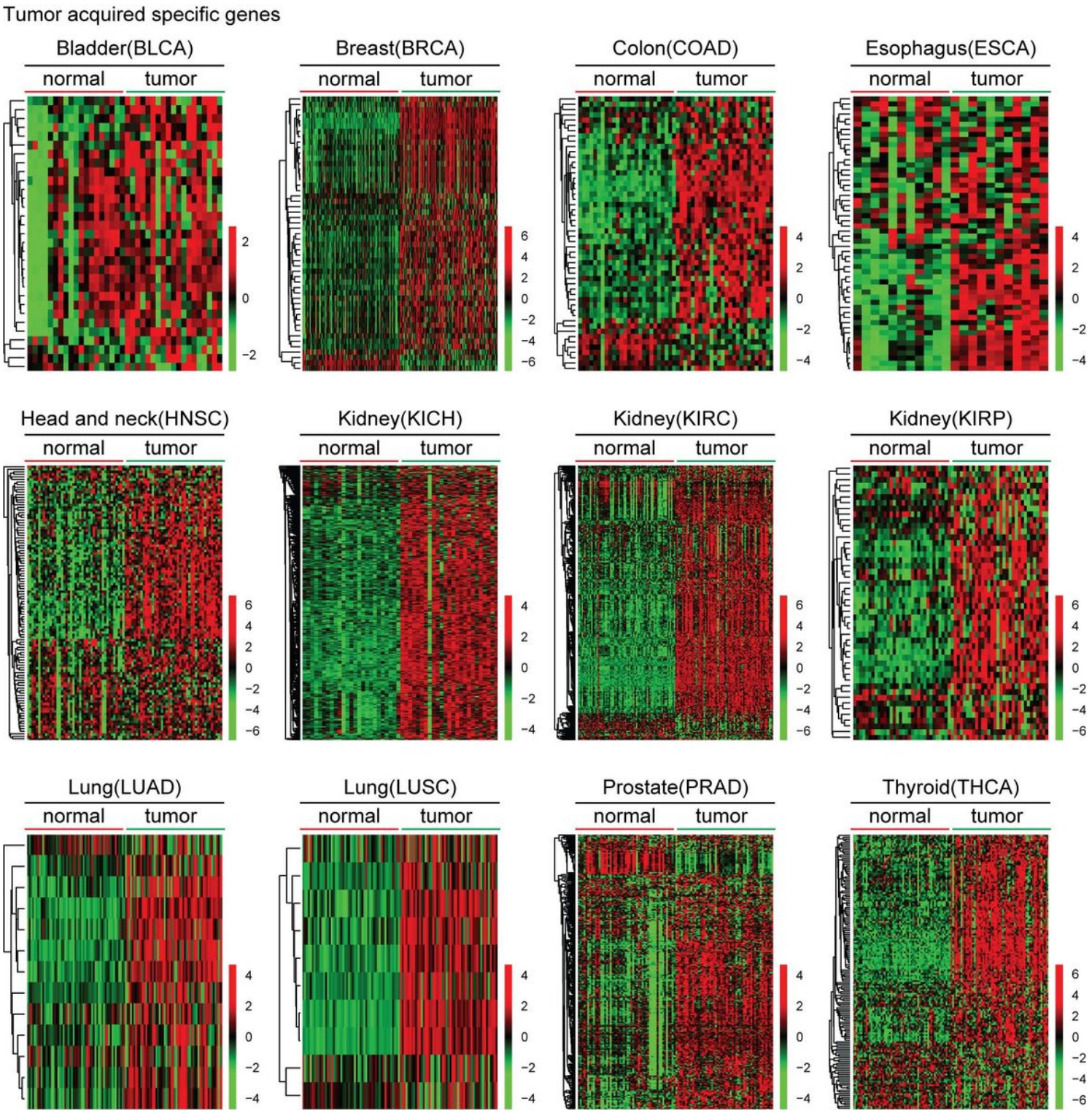

Figure 6

The tumor acquired specific genes are increased in the tumor samples. Heatmaps showed the expressions (log2 count) of tumor acquired specific genes in normal and tumor samples in BLCA, BRCA, COAD, ESCA, HNSC, KICH, KIRC, KIRP, LUAD, LUSC, PRAD and THCA tumor types.

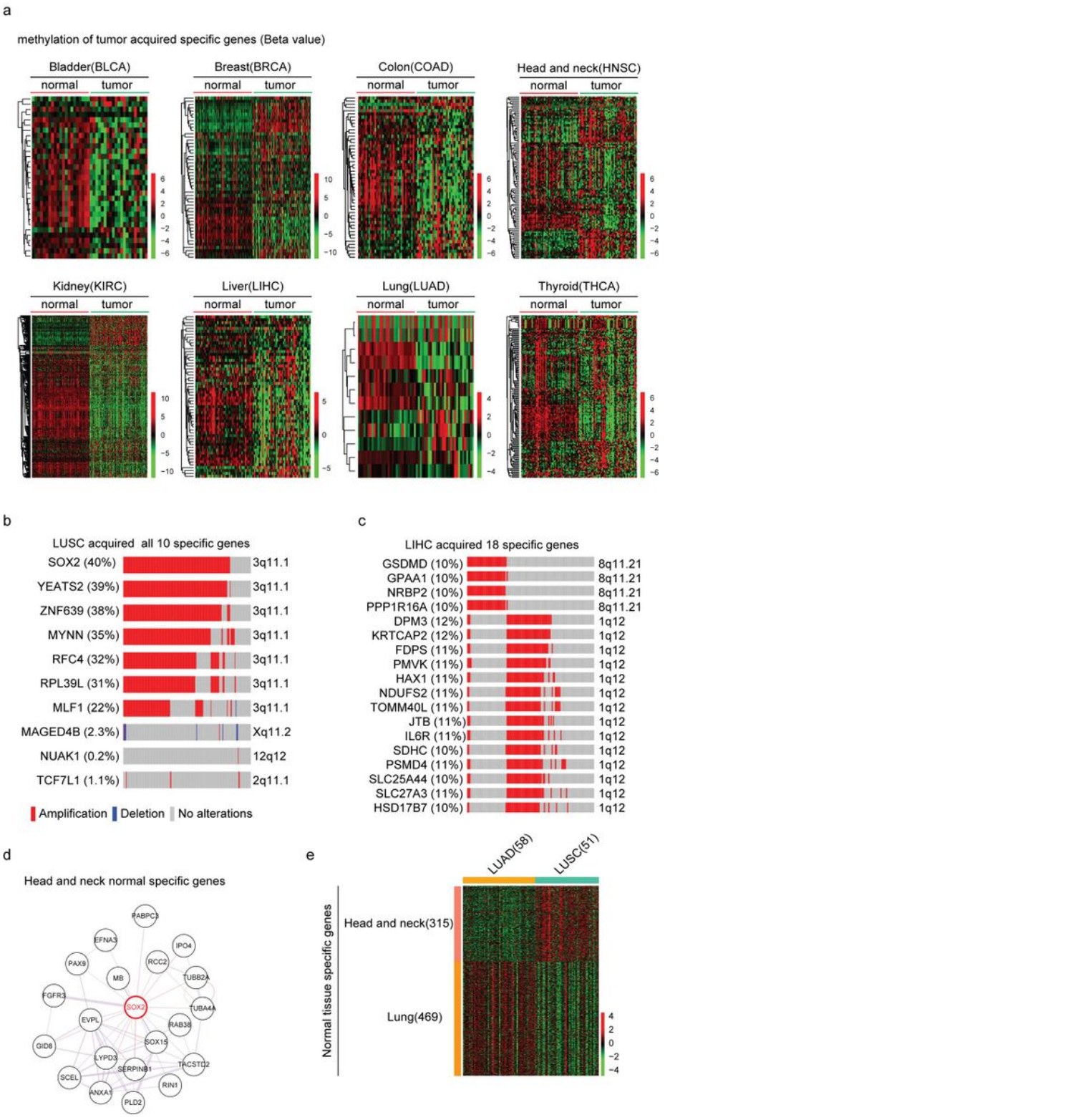

Figure 7

The expressions of tumor acquired specific genes are increased by hypo-DNA methylation and DNA amplification. (a) Heatmaps showed the methylation (Beta value) of tumor acquired specific genes in normal and tumor samples in each tumor type. Hyper-methylated (red) and hypo-methylated (green) genes were delineated. (b) Oncoprints demonstrated the tumor acquired specific genes in LUSC. Each bar represented one patient. (c) Oncoprints demonstrated the tumor acquired specific genes in LIHC. (d)

Normal tissue specific core transcription factors mediated regulatory gene networks were created by cytoscape. Genes connected with SOX2 from normal head and neck specific genes were constructed. (e) Normal lung, head and neck specific genes were demonstrated in LUAD and LUSC tumor types through heatmap.

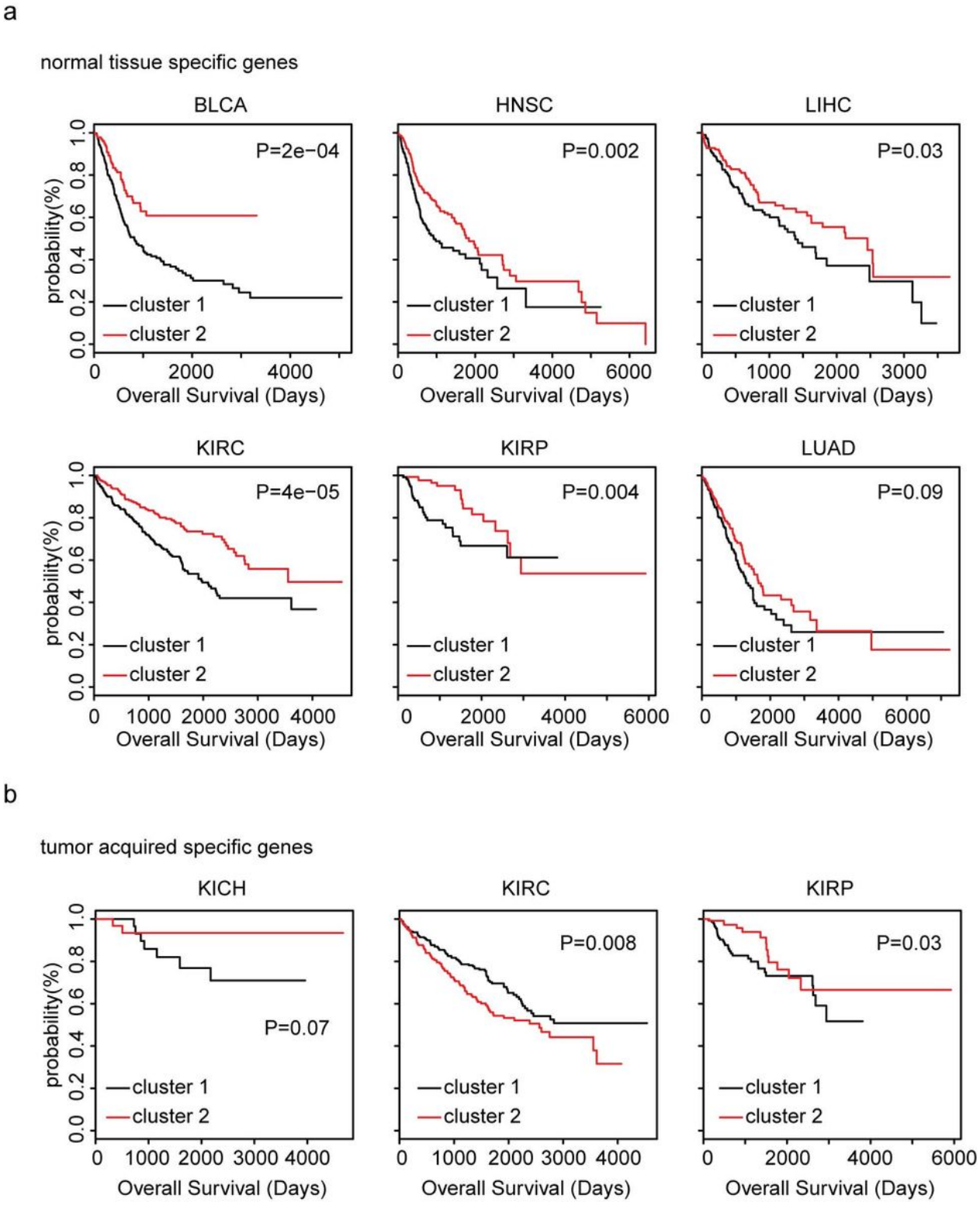

Figure 8

The normal tissue specific genes and the tumor acquired specific gene expression are associated with the tumor outcomes. (a) Associations between normal tissue specific genes and tumor overall survival were studied from TCGA expression dataset. Tumor patients were divided into two groups by the expression of tissue specific genes. Kaplan-Meier analysis was used to compare the overall survival of two groups. Log-rank test showed overall p-value. (b) Association between tumor acquired specific genes and overall survival was studied from TCGA expression dataset.